

68 FIG. 3 AU - MARTINEZ R AU - EDWARDS CA TI - EXPRESSION, PURIFICATION AND FUNCTIONAL CHARACTERIZATION OF THE **DNA-BINDING DOMAIN OF THE HERPES SIMPLEX VIRUS TYPE 1 UL9 PROTEIN.** 70 **RN - EC 3.4.21.5 (THROMBIN)** RN - 0 (VIRAL PROTEINS) RN - 115004-77-8- (HERPES SIMPLEX **VIRUS TYPE 1 PROTEIN UL9)** RN - 9007-49-2 (DNA) **THROMBIN** VIRAL PROTEINS **HERPES SIMPLEX VIRUS TYPE 1 PROTEIN UL9** DNA 72 **FILTER** WORDS: THROMBIN VIRAL **HERPES SIMPLEX VIRUS TYPE 1 PROTEINS PROTEIN UL9** DNA 76 74 **CO-OCCURRENCE COUNTS INFERENCE DATABASE** ID1 ID2 COUNT ID NAME 1 2 12 1 THROMBIN 2 HERPES SIMPLEX... 1 3 4 44 3 DNA 12 12 78 80 80' 44' 44 5 5 **INFERENCE: HERPES SIMPLEX VIRUS TYPE 1** PROTEIN UL9 INTERACTS WITH DNA 82

START

86

CREATING A CONNECTION NETWORK FROM AN INFERENCE DATABASE, WHERE THE CONNECTION NETWORK INCLUDES TWO OR MORE NODES CONNECTED BY ONE OR MORE ARCS, WHERE THE ONE OR MORE ARCS REPRESENTS CO-OCCURRENCES BETWEEN CHEMICAL OR BIOLOGICAL MOLECULES, WHERE THE INFERENCE DATABASE INCLUDES DATABASE RECORDS WITH ONE OR MORE INFERENCE ASSOCIATIONS

88

APPLY ONE OR MORE ANALYSIS METHODS TO THE CONNECTION NETWORK TO DETERMINE ANY TRIVIAL INFERENCE ASSOCIATIONS

90

DELETE DATABASE RECORDS FROM THE INFERENCE DATABASE DETERMINED TO INCLUDE TRIVIAL INFERENCE ASSOCIATIONS, THEREBY IMPROVING INFERENCE KNOWLEDGE IN THE INFERENCE DATABASE

92

START

EXTRACT TWO OR MORE CHEMICAL OR
BIOLOGICAL MOLECULES NAMES FROM A
DATABASE RECORD FROM AN INFERENCE
DATABASE FOR A FIRST CHEMICAL OR
BIOLOGICAL MOLECULE-A AND A SECOND
CHEMICAL OR BIOLOGICAL MOLECULE-B. THE
INFERENCE DATABASE INCLUDES A PLURALITY
OF INFERENCE DATABASE RECORDS CREATED
FROM AN INDEXED LITERATURE DATABASE

DETERMINE A LIKELIHOOD STATISTIC FOR A CO-OCCURRENCE BETWEEN A FIRST CHEMICAL OR BIOLOGICAL MOLECULE NAME-A AND A SECOND CHEMICAL OR BIOLOGICAL MOLECULE NAME-B EXTRACTED FROM THE DATABASE RECORD

APPLY THE LIKELIHOOD STATISTIC TO
DETERMINE IF THE CO-OCCURRENCE BETWEEN
THE FIRST CHEMICAL OR BIOLOGICAL
MOLECULE-A AND THE SECOND CHEMICAL OR
BIOLOGICAL MOLECULE-B IS A NON-TRIVIAL COOCCURRENCE REFLECTING PHYSICO-CHEMICAL
INTERACTIONS

END

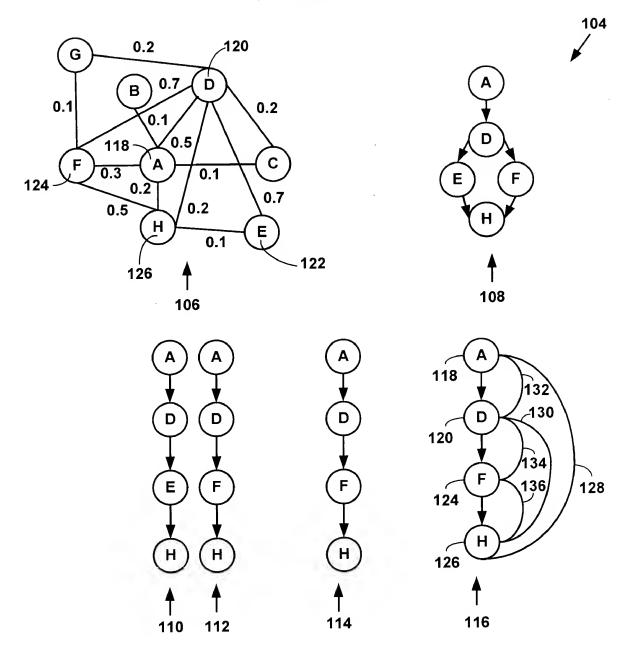
100

96

98

102

FIG. 6



START

140

SELECT A TARGET NODE FROM A FIRST LIST OF NODES CONNECTED BY A PLURALITY OF ARCS IN A CONNECTION NETWORK, WHERE THE CONNECTION NETWORK INCLUDES ONE OR MORE NODES REPRESENTING ONE OR MORE CHEMICAL OR BIOLOGICAL MOLECULES NAMES AND ONE OR MORE ARCS CONNECTING THE ONE OR MORE NODES IN A PRE-DETERMINED ORDER, AND WHERE THE ONE OR MORE ARCS REPRESENT CO-OCCURRENCE VALUES OF PHYSICO-CHEMICAL INTERACTIONS BETWEEN CHEMICAL OR BIOLOGICAL MOLECULES

142

CREATE A SECOND LIST OF NODES BY
CONSIDERING SIMULTANEOUSLY ONE OR MORE
OTHER NODES THAT ARE NEIGHBORS OF THE
TARGET NODE AS WELL AS NEIGHBORS OF THE
OTHER NODES IN THE PRE-DETERMINED ORDER
IN THE CONNECTION NETWORK

144

SELECT A NEXT NODE FROM THE SECOND LIST OF NODES USING THE CO-OCCURRENCE VALUES, WHERE THE NEXT NODE IS A MOST LIKELY NEXT NODE AFTER THE TARGET NODE IN THE PREDETERMINED ORDER FOR THE CONNECTION NETWORK BASED ON THE CO-OCCURRENCE VALUES

146

148 **START** SELECT A POSITION IN A CONNECTION NETWORK FOR AN UNKNOWN TARGET NODE FROM A FIRST LIST OF NODES CONNECTED BY ONE OR MORE ARCS. THE CONNECTION NETWORK INCLUDES ONE OR MORE NODES REPRESENTING ONE OR MORE CHEMICAL OR BIOLOGICAL MOLECULES NAMES AND ONE OR MORE ARCS CONNECTING THE ONE OR MORE NODES IN A PRE-DETERMINED 150 ORDER. THE ONE OR MORE ARCS REPRESENT CO-OCCURRENCE VALUES OF PHYSICO-CHEMICAL INTERACTIONS BETWEEN CHEMICAL OR BIOLOGICAL MOLECULES. DETERMINE A SECOND LIST OF NODES PRIOR TO THE POSITION OF THE UNKNOWN TARGET NODE IN THE CONNECTION NETWORK 152 **DETERMINE A THIRD LIST OF NODES** SUBSEQUENT TO THE POSITION OF UNKNOWN TARGET NODE IN THE CONNECTION NETWORK 154 DETERMINE A FOURTH LIST OF NODES INCLUDED IN BOTH THE SECOND LIST OF NODES AND THE THIRD LIST OF NODES 156 DETERMINE AN IDENTITY FOR THE UNKNOWN TARGET NODE FROM THE FOURTH LIST OF NODES **USING A LIKELIHOOD STATISTIC** 158

START

160

CONSTRUCT A CONNECTION NETWORK USING ONE OR MORE DATABASE RECORDS FROM AN INFERENCE DATABASE. THE CONNECTION NETWORK INCLUDES A ONE OR MORE NODES FOR CHEMICAL OR BIOLOGICAL MOLECULES AND BIOLOGICAL PROCESSES FOUND TO CO-OCCUR ONE OR MORE TIMES. THE ONE OR MORE NODES ARE CONNECTED BY ONE OR MORE ARCS IN A PRE-DETERMINED ORDER. THE INFERENCE DATABASE WAS CREATED FROM CHEMICAL OR BIOLOGICAL MOLECULE AND BIOLOGICAL PROCESS INFORMATION EXTRACTED FROM A STRUCTURED LITERATURE DATABASE.

162

APPLY ONE OR MORE LIKELIHOOD STATISTIC
ANALYSIS METHODS TO THE CONNECTION
NETWORK TO DETERMINE POSSIBLE INFERENCES
BETWEEN THE CHEMICAL OR BIOLOGICAL
MOLECULES AND A BIOLOGICAL PROCESS

164

GENERATE AUTOMATICALLY ONE OR MORE BIOLOGICAL INFERENCES BETWEEN CHEMICAL OR BIOLOGICAL MOLECULES AND A BIOLOGICAL PROCESS USING RESULTS FROM THE LIKELIHOOD STATISTIC ANALYSIS METHODS

166